Original Article
Exploring the molecular mechanism, biomarkers and prognostic values of HCC based on gene expression microarray

Ding Li1*, Fu-Biao Ni1*, Heng-Jie Tang1, Chong-Chu Huang1, Cheng-Yang Liu2, Ji-Cai Chen1

1Department of General Surgery, The First Affiliated Hospital of Wenzhou Medical University, Wenzhou, Zhejiang, China; 2Department of Surgery, School of Medicine, University of Pennsylvania, Philadelphia, PA, USA. *Equal contributors.

Received March 6, 2019; Accepted May 13, 2019; Epub August 15, 2019; Published August 30, 2019

Abstract: Liver cancer is a common malignant neoplasm worldwide, causing high morbidity and mortality globally. The molecular mechanisms of hepatocarcinogenesis remains unclear. The goal of our study is to elucidate the mechanism which could improve the prognosis of liver cancer. GSE121248 was downloaded from the GEO database, which is a gene expression profile data including 70 tumor samples and 37 adjacent normal samples from hepatocellular carcinoma (HCC). The differentially expressed genes (DEGs) between cancer tissues and normal tissues were screened. Subsequently, the enriched GO terms, KEGG pathways and Database for Annotation, Visualization and Integrated Discovery (DAVID) were analyzed by on-line tools. Finally, STRING database and Cytoscape software were used to construct protein-protein interactions (PPI) network and genes with high degree. Kaplan-Meier plotter (KM plotter) was used to explore the predictive prognostic value of gene expression to overall survival (OS). The results showed that there were 202 DEGs between the case samples and control samples, including 59 up-regulated genes and 143 down-regulated genes in HCC tissue samples. The study obtained a total of 132 enriched GO terms and 16 KEGG pathways of those genes and most of the enriched GO terms that were correlated to HCC were also included in this research. The patients with high expression of 5 core genes had poor OS according to online database. Using comprehensive bioinformatics analyses, our study attempts to identify DEGs and find potential biomarkers to predict the occurrence and development of HCC.

Keywords: Molecular mechanism, biomarkers, prognostic values, HCC, gene expression microarray

Background

Primary liver cancer (namely hepatocellular carcinoma, HCC) is one of the fifth most common cancer which takes up the third top cause of cancer mortality [1]. In 2016, it afflicted in excess of 1 million people and caused 800,000 deaths globally [2]. HCC has interesting epidemiologic features and a couple of well-substantiated environmental potentially preventable risk factors have been confirmed for the disease [3]. Despite active research [1, 4, 5], the molecular mechanisms that induce HCC remain unclear.

Gene expression profiles measure the expression of thousands of genes to build a global picture of cell function. Gene expression profiling may become an important diagnostic test [6, 7]. Such new technique may have positive influence in improving our knowledge of carcinogenesis and facilitate screening and early detection of diseases.

In our study, we analyzed gene expression profiles (GSE121248) downloaded from GEO database, to extract DEGs between liver cancer and normal samples. Moreover, KEGG pathways of those genes and most of the enriched GO terms that were correlated to HCC were also included in this research. In summary, our result may advance the knowledge of HCC and explore potential therapeutic targets for further studies.
Materials and methods

Gene expression microarray data

In this research, the gene expression microarray data set GSE121248 was downloaded from GEO database (GEO, http://www.ncbi.nlm.nih.gov/geo/), including 70 tumor samples from HCC patient and 37 adjacent normal samples [8]. We used k-nearest neighbors algorithm (k-NN) for classification and regression in pattern recognition [9]. In addition, a supplement work was conducted for other genes with similar expression profiles as those with deletion value.

Differentially expressed genes (DEGs)

In the preprocessing section, the unwanted noise of the original microarray data was screened out. The original data and background correction were dealt with by the Affy [10] package in R. The differentially expressed mRNAs between tumor tissue and normal tissue samples were inspected using the Limma [11] package with the following criteria: |Log2 (fold change)|>1.5 and false discovery rate <0.05. Morpheus is an on-line tool used to deal with heatmap (https://software.broadinstitute.org/morpheus).

Functional and pathway enrichment analysis

The gene product function is supplied from The Gene Ontology website [12] (GO, http://www.genneontology.org), a community-based bioinformatics resource using ontologies to represent biological knowledge. GO provides a framework and a suit of concepts for describing the functions of gene products from all living organisms. KEGG (Kyoto Encyclopedia of Genes and Genomes, http://www.kegg.jp/ or http://www.genome.jp/kegg) is used for interpretation of high-throughput data and other biological genome sequences [13]. DAVID [14] (https://david.ncifcrf.gov) is also a bioinformatic resources, consisted of analytic tools and biological knowledge base, aimed at organically extracting biological activity from huge gene protein statement. GO terms and KEGG pathway analyses were accomplished with DAVID to recognize DEGs. P<0.05 was executed as the cut-off value.

Protein-protein interactions (PPI) network construction and analysis of modules

The STRING database (http://string-db.org) is a web resource aimed to predict protein-protein interactions and provide critical assessment, which includes physical and functional associations [15]. Cytoscape [16] is an open-source software platform for visualizing biomedical networks exploration and offering researchers utility and coactive visualization interface. The tools is used for exploring interconnections of complex biological networks supported by varying annotation and experimental results, therefore accelerating research works such as forecasting gene function and reconstructing pathways. The STRING was used to evaluate PPI network mapped DEGs and Cytoscape for visualization. The combined score of greater than 0.15 was executed as the cut-off value. Finally the cut-off criterion for screening hub gene was defined as node degree 10.

Exploring overall survival

‘Kaplan-Meier plotter’ (KM plotter), online database, was used to explore the predictive prognostic value of gene expression to overall survival (OS) in different clinical data, such as clinical grades, clinical stages, gender and race in liver cancer patients.

Statistical analysis

Matrix data were analyzed by one-way ANOVA, and statistical significance was determined as p<0.05. The results were compared between normal and tumor groups with Mann-Whitney Rank Sum Test.

Results

Differentially expresses gene (DEGs)

There were 202 DEGs between tumor samples and control samples. Comparing with adjunct tissue samples, 59 up-regulated and 143 down-regulated genes were selected in case tissue samples. The heatmap of the DEGs was shown in Figure 1.

GO enrichment terms and KEGG pathway analysis

The study obtained totally of 132 GO enrichment terms and 16 KEGG pathways. The first 10 GO enrichment terms of the DEGs following P value were shown (Table 1). Table 1 indicated clearly that the mainly GO enrichment term was the cellular components (CC), such as extracellular region, organelle membrane and extracellular space. In addition, epoxygenase P450
Figure 1. Heatmap of 202 DEGs. 70 tumor cases and 37 normal group had 202 DEGs, including 59 up-regulated genes and 143 down-regulated genes in HCC samples.

pathway and oxidation-reduction process were also the enriched GO terms related to the cell biological process (BP). Arachidonic acid epoxygenase activity, oxidoreductase activity, iron
Bioinformation analysis of HCC

Table 1. Top 10 enriched GO terms are sorted in ascending order by P value

<table>
<thead>
<tr>
<th>Category</th>
<th>Term</th>
<th>Gene Number</th>
<th>P Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>CC</td>
<td>G:0005576~extracellular region</td>
<td>47</td>
<td>7.04E-11</td>
</tr>
<tr>
<td>BP</td>
<td>G:0019373~epoxygenase P450 pathway</td>
<td>8</td>
<td>4.33E-10</td>
</tr>
<tr>
<td>CC</td>
<td>G:0031090~organelle membrane</td>
<td>12</td>
<td>1.25E-09</td>
</tr>
<tr>
<td>MF</td>
<td>G:0008392~arachidonic acid epoxygenase activity</td>
<td>7</td>
<td>5.10E-09</td>
</tr>
<tr>
<td>MF</td>
<td>G:0016705~oxidoreductase activity, acting on paired donors, with incorporation or reduction of molecular oxygen</td>
<td>10</td>
<td>6.28E-09</td>
</tr>
<tr>
<td>MF</td>
<td>G:0005506~iron ion binding</td>
<td>13</td>
<td>6.16E-08</td>
</tr>
<tr>
<td>MF</td>
<td>G:0004497~monooxygenase activity</td>
<td>9</td>
<td>1.34E-07</td>
</tr>
<tr>
<td>MF</td>
<td>G:0020037~heme binding</td>
<td>12</td>
<td>1.75E-07</td>
</tr>
<tr>
<td>CC</td>
<td>G:0005615~extracellular space</td>
<td>36</td>
<td>2.19E-07</td>
</tr>
<tr>
<td>BP</td>
<td>G:00055114~oxidation-reduction process</td>
<td>23</td>
<td>4.31E-07</td>
</tr>
</tbody>
</table>

Table 2. The top 10 KEGG pathways in DEGs

<table>
<thead>
<tr>
<th>Category</th>
<th>Pathway Name</th>
<th>Gene Number</th>
<th>P Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>KEGG_PATHWAY</td>
<td>hsa04610: Complement and coagulation cascades</td>
<td>8</td>
<td>3.84E-05</td>
</tr>
<tr>
<td>KEGG_PATHWAY</td>
<td>hsa00830: Retinol metabolism</td>
<td>7</td>
<td>2.20E-04</td>
</tr>
<tr>
<td>KEGG_PATHWAY</td>
<td>hsa05020: Prion diseases</td>
<td>5</td>
<td>1.11E-03</td>
</tr>
<tr>
<td>KEGG_PATHWAY</td>
<td>hsa01100: Metabolic pathways</td>
<td>30</td>
<td>1.26E-03</td>
</tr>
<tr>
<td>KEGG_PATHWAY</td>
<td>hsa00232: Caffeine metabolism</td>
<td>3</td>
<td>1.80E-03</td>
</tr>
<tr>
<td>KEGG_PATHWAY</td>
<td>hsa00380: Tryptophan metabolism</td>
<td>5</td>
<td>2.05E-03</td>
</tr>
<tr>
<td>KEGG_PATHWAY</td>
<td>hsa05204: Chemical carcinogenesis</td>
<td>6</td>
<td>4.59E-03</td>
</tr>
<tr>
<td>KEGG_PATHWAY</td>
<td>hsa04115: p53 signaling pathway</td>
<td>5</td>
<td>1.30E-02</td>
</tr>
<tr>
<td>KEGG_PATHWAY</td>
<td>hsa00982: Drug metabolism-cytochrome P450</td>
<td>5</td>
<td>1.37E-02</td>
</tr>
<tr>
<td>KEGG_PATHWAY</td>
<td>hsa04978: Mineral absorption</td>
<td>4</td>
<td>2.19E-02</td>
</tr>
<tr>
<td>KEGG_PATHWAY</td>
<td>hsa05323: Rheumatoid arthritis</td>
<td>5</td>
<td>3.19E-02</td>
</tr>
<tr>
<td>KEGG_PATHWAY</td>
<td>hsa00140: Steroid hormone biosynthesis</td>
<td>4</td>
<td>4.44E-02</td>
</tr>
<tr>
<td>KEGG_PATHWAY</td>
<td>hsa04060: Cytokine-cytokine receptor interaction</td>
<td>8</td>
<td>4.79E-02</td>
</tr>
</tbody>
</table>

ion binding, monooxygenase activity and heme binding composed the molecular function (MF) of the cell.

The KEGG pathways were demonstrated in Table 2. The top 10 KEGG enriched pathways were mostly related to metabolism, such as retinol metabolism, metabolic pathways, caffeine metabolism, tryptophan metabolism and drug metabolism-cytochrome P450. Furthermore, the other pathways may also influence on the progression of cancer via some biological process, such as the complement and coagulation cascades, Prion diseases, chemical carcinogenesis, p53 signaling pathway and mineral absorption.

**PPI network and core genes in network**

The PPI (Figure 2) network included 190 nodes (DEGs) and 756 edges (interactions) between the DEGs. The core genes, had higher degree in the PPI network, might had a stronger correlation with liver cancer. Table 3 has shown the cores genes' solid degree. Among those core genes, CDC20 had 33 degrees and 21 gene's node degree were beyond 30.

**Overall survival based on core genes**

Firstly, the prognostic value of BUB1B was accessed in the database. The RNA-seq ID is 701 for BUB1B. OS curves were plotted for all liver cancer patients (n=364) (Figure 3A). BUB1B has related to poor OS in HCC patients, HR=2.01 (1.42-2.86), P=6.6e-05. Next, the prognostic significance of CCNB1 was evaluated in the database. The desired RNA-seq ID is 891. CCNB1 revealed a remarkable correlation with worse OS among liver cancer patients, HR=2.34 (1.55-3.54), P=3.4e-05 (Figure 3B). Similarly, EZH2 (RNA-seq ID 2146) exhibited a
Figure 2. The PPI network of 190 nodes and 756 edges. The 47 red nodes had higher expression in the tumor samples; whereas the 103 blue nodes have lower expression values.
Bioinformation analysis of HCC

**Table 3.** The cores genes and solid degree according to the PPI network

<table>
<thead>
<tr>
<th>Gene symbol</th>
<th>Node degree</th>
<th>Gene symbol</th>
<th>Node degree</th>
<th>Gene symbol</th>
<th>Node degree</th>
<th>Gene symbol</th>
<th>Node degree</th>
</tr>
</thead>
<tbody>
<tr>
<td>CYP26A1</td>
<td>10</td>
<td>C8A</td>
<td>15</td>
<td>CDKN3</td>
<td>30</td>
<td>ZWINT</td>
<td>31</td>
</tr>
<tr>
<td>CYP2B6</td>
<td>10</td>
<td>F9</td>
<td>16</td>
<td>HMMR</td>
<td>30</td>
<td>BUB1B</td>
<td>32</td>
</tr>
<tr>
<td>ALDH8A1</td>
<td>11</td>
<td>IGF1</td>
<td>16</td>
<td>KIF4A</td>
<td>30</td>
<td>CNB1</td>
<td>32</td>
</tr>
<tr>
<td>C6</td>
<td>11</td>
<td>SPP1</td>
<td>17</td>
<td>PTTG1</td>
<td>30</td>
<td>EZH2</td>
<td>32</td>
</tr>
<tr>
<td>CYP1A2</td>
<td>11</td>
<td>GINS1</td>
<td>20</td>
<td>ASPM</td>
<td>31</td>
<td>NUSAP1</td>
<td>32</td>
</tr>
<tr>
<td>CYP2C9</td>
<td>11</td>
<td>CENPW</td>
<td>22</td>
<td>DLGAP5</td>
<td>31</td>
<td>CDC20</td>
<td>33</td>
</tr>
<tr>
<td>EGR1</td>
<td>11</td>
<td>E2F8</td>
<td>22</td>
<td>DTL</td>
<td>31</td>
<td></td>
<td></td>
</tr>
<tr>
<td>HGFAC</td>
<td>11</td>
<td>CENPK</td>
<td>23</td>
<td>KIF20A</td>
<td>31</td>
<td></td>
<td></td>
</tr>
<tr>
<td>FOS</td>
<td>12</td>
<td>UHRF1</td>
<td>23</td>
<td>NCPG</td>
<td>31</td>
<td></td>
<td></td>
</tr>
<tr>
<td>KLKB1</td>
<td>12</td>
<td>FAM83D</td>
<td>25</td>
<td>NDC80</td>
<td>31</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SERPINE1</td>
<td>12</td>
<td>ANLN</td>
<td>27</td>
<td>PBK</td>
<td>31</td>
<td></td>
<td></td>
</tr>
<tr>
<td>MBL2</td>
<td>13</td>
<td>UBE2T</td>
<td>27</td>
<td>RAD51AP1</td>
<td>31</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CCL2</td>
<td>14</td>
<td>RACGAP1</td>
<td>28</td>
<td>RRM2</td>
<td>31</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CXCL12</td>
<td>14</td>
<td>NUF2</td>
<td>29</td>
<td>TOP2A</td>
<td>31</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SPP2</td>
<td>14</td>
<td>PRC1</td>
<td>29</td>
<td>TTK</td>
<td>31</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

worse correlation with OS within liver cancer patients, HR=2.23 (1.56-3.19), P=6.8e-06 (Figure 3C). The prognostic value of NUSAP1 was also determined in the database, RNA-seq ID 51203. A strong association was identified between NUSAP1 expression and poor OS for liver cancer, HR=1.67 (1.17-2.4), P=0.0046 (Figure 3D). Figure 3E demonstrated the prognostic value of CDC20 in the database. The RNA-seq ID is 991. CDC20 was significantly correlated with poor OS for liver cancer patients, HR=2.49 (1.72-3.59), P=5.1e-07.

**Discussion**

HCC’s mortality has steadily increased in the last few years. It is now the fifth most common malignant tumor worldwide [17]. Hepatitis viruses, gene mutations, cell damage, alcoholic liver diseases and aflatoxin poisoning have all been identified as risk factors for HCC [18]. As a potential diagnostic tool, tumor markers have been widely used in the early diagnosis of HCC. Alejandro Forner claimed that α-Fetoprotein was a brilliant star for HCC diagnosis [19]. Moreover, severe alpha-1-antitrypsin deficiency (AATD) played a key role in the development of liver disease [20]. However, further study on the mechanism of hepatocellular carcinoma is necessary for early diagnosis and optimal treatment.

In this work, a high-throughput method for genome-wide gene expression analysis was used to screen 202 DEGs from HCC tissues and normal liver tissue, including 59 up-regulated genes and 143 down-regulated genes in tumor samples.

Enrichment GO analyses identified significant ontology categories including epoxygenase P450 pathway and oxidation-reduction process. The human liver microsome metabolizes arachidonic acid in NADPH and produces epoxyeicosapentaenoic acid and its hydrated metabolite dihydroxyeicosapentaenoic acid as the main reaction product [21]. These bioactive eicosanoids play a role in maintaining homeostasis in the liver. Under normal and pathophysiological conditions, human P450 cyclooxygenase (also known as CYP) and its derivatives arachidonic acid metabolites may be expressed differently, thus affecting regulation of vascular function [22]. In one study, it was found that CYP2C9 was highly expressed in patients with esophageal adenocarcinoma (including patients with early tumor stage and highly differentiated tumors). Additionally, selective inhibition of CYP2C9 was shown to reduce the proliferation of low tumor cells in in vitro experiments [23].

KEGG pathway analyzed DEGs and 13 pathways were screened out, such as retinol metabolism, tryptophan metabolism and drug metabolism-cytochrome P450, complement and coagulation cascades. Previous studies have covered that majority of these pathways were comprised in cancer progression. In human cancer cell lines of skin, oral cavity, kidney and breast, retinol acyltransferase (LRAT) was reduced compared with the normal counterparts, thus implicating aberrant retinoid metabolism in carcinogenesis [24]. Cytochrome P450 are hepatic enzymes that may activate some procarcinogens. Studies have shown that CYP1A1 increases the risk of peripheral adenocarcinoma type lung cancer [25]. And its polymorphism might be an important factor in the optimal use of selected anticancer drugs for cancer treatment [26].
Figure 3. The prognostic HRs value of BUB1B, CCNB1, EZH2, NUSAP1 and CDC20 in liver cancer in www.kmplot.com. (A: BUB1B, HR=2.01 (1.42-2.86), P=6.6e-05. B: CCNB1, HR=2.34 (1.55-3.54), P=3.4e-05. C: EZH2, HR=2.23 (1.56-3.19), P=6.8e-06. D: NUSAP1, HR=1.67 (1.17-2.4), P=0.0046. E: CDC20, HR=2.49 (1.72-3.59), P=5.1e-07).
The complement system is responsible for killing bacteria that infect the host. The secretion of proinflammatory mediators and ingestion of opsonized particles were led by complement activation [27]. On the other hand, coagulation activation occurs through two pathways, one exogenous and the other through an internal (contact activation) pathway that seems to be involved in inflammatory processes [28, 29]. In the occurrence and progression of tumor, tumor-promoting inflammation plays an important role [30, 31]. The connection between inflammation and cancer can be made both externally and internally. The external pathway increases the risk of inflammation driving cancer, while the internal pathway is driven by genes for inflammation and tumorigenesis [32].

In addition, PPI suggests that the top 5 core genes were BUB1B, CCNB1, EZH2, NUSAP1 and CDC20, which might serve as potential targets for therapy. In normal cells, BUB1B (coding BUBR1) has been shown to prevent duplication of chromosome segregation, but its role has been controversial in cancer pathogenesis. BUB1B has overexpression in prostate cancer [33], adenomatous polyposis coli [34], HCC [35] and other cancer [36, 37], which is consistent with our data; while low expression of BUB1B contributes to embryonal rhabdomyosarcoma [38], colorectal cancer [39] and other cancers [40, 41]. CCNB1 (Cyclin B1), tumor antigen, is overexpressed in many cancers [42]. The autoimmune response of CCNB1 in HCC may include the aberration of CCNB1 regulation leading to changes in product or its expression resulting in immune stimulation [43]. Wu suggest that CCNB1 may be a key target protein of human HCC cell line Lnc00312 [44]. Enhancer of zeste homolog 2 (EZH2) is a histone-lysine N-methyltransferase enzyme, which participates in histone methylation and transcriptional repression [45]. In some reports, inhibition of EZH2 function shrunk malignant tumors because the tumor suppressor genes were not inhibited by EZH2 [46]. In HCC, EZH2 partially inhibits the immune response and plays a carcinogenic role [47]. Nucleolar and spindle-associated protein 1 (NUSAP1) maintains normal cellular division and participates in regulating spindle assembly through microtubule-binding and DNA-binding domains, which is an important regulator of mitosis and cell proliferation [48, 49]. Previous study indicated that the expression of NUSAP1 at the margin of liver cancer surgery was closely related to early postoperative recurrence and can be used as an index to predict early recurrence of HCC [50]. As a cell cycle regulating kinase, CDC20, an essential cell cycle regulator, is necessary to complete mitosis. Some studies have reported that CDC20 plays a key role in gastric cancer [51] and HCC [52]. CDC20 may potentially be used as a biomarker and therapeutic target in HCC [53].

Finally, our results showed that high expression of the top 5 core genes had worsening effect on the prognosis of liver cancer patients. Zhuang [54] reported that BUB1B, CCNB1 and CDC20 could serve as predictive biomarkers for HCC, and they demonstrated that high expression of those genes was related to worse survival. As for NUSAP1, it was a valuable prognostic factor for hepatic carcinoma. Low NUSAP1 expression patients had better survival rate than high expression patients at both 6 months and 12 months (89.3% VS 33.3%, 53.6% VS 17.9%) [55]. It was reported that precise anti-tumor drugs can be developed because EZH2 promotes the occurrence and development of tumors [56]. This has the potential to improve survival in cancer patients.

**Conclusion**

Our study screened DEGs and identified potential biomarkers to forecast the occurrence and development of HCC. A total of 202 DEGs were screened including BUB1B, CCNB1, EZH2, NUSAP1 and CDC20. The many functional partnerships and PPI are core of cellular processing and their classified characterization helps to deal with context in molecular system biology. Survival analysis identified 5 core genes as potential therapeutic targets in the management of HCC. However, our study lacks in vivo and in vitro validation. For future studies, the results of these bioinformatics analyses can be verified by experiments, such as Western Blot and qRT-PCR.

**Acknowledgements**

Thanks are due to Dr. Xian-Wei Zhang (The First Affiliated Hospital of Wenzhou Medical University, Wenzhou, Zhejiang, 325000, China) for assistance with the R studio and to Dr. Zhi-Yang Zhou (The First Affiliated Hospital of Wenzhou
References


Bioinformation analysis of HCC


